Partner File Processing
Address Updates

Census Bureau
Community Addressing Conference
Leesburg, VA

April 18, 2013

Overview

- Two major differences between the GSS-I Address Partnership Program and the Local Update of Census Addresses (LUCA)

  1) Accepting data submissions in native format rather than pre-defined template
  2) Office verification in lieu of field verification

  The result is a very different processing flow…

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU

File Preparation Stage

ETL (Extract/Transform/Load)

Preprocessing Edits

Match and Geocode

Create Evaluation Package

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU

Step 1: Extract/Transform/Load (ETL)

- Crosswalk the partner data to a standard format, load into an Oracle table for subsequent processing

- If necessary, combine separate layers to derive address class codes, address coordinates

- Perform parsing and concatenation, field and value mapping using FME (Feature Management Engine)

- Process includes data conversion (where applicable) to meet FGDC Address Standard

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU

4

Step 2: Preprocessing Edits

- Check and/or correct for data issues
- Examples:
  - Illegal Characters
  - '0' House Numbers
  - Truncated ZIP Codes

- Run "Header Coding" to correct erroneous or derive missing ZIP Codes

Step 3: Geocode and Match to MTDB

- Using provided or derived address coordinates, determine census blocks

- Geocode to census block using TIGER Address Ranges

- Match addresses to the Master Address File and assign outcome codes

United States™ **Census** Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU

Step 4: Create Evaluation Work Package

- Group addresses into evaluation categories

- Create sample sets for review

- Load the matched, geocoded data into ESRI environment

- Add additional data layers (tabulation blocks, linear features, coverage analysis layer)

Address Data Content Guidelines 1

In order to perform an update to the Census Bureau's Master Address File (MAF), we must first be able to perform a match between addresses in the partner file and addresses already existing in the MAF – otherwise, we risk adding duplicate records to the MAF. Therefore:

A) In order to perform a match to existing MAF addresses the submitted record must include:

- Complete Address Number
- Complete Street Name

    <u>and</u> AT LEAST ONE OF THE FOLLOWING:

- Address Coordinate
- ZIP Code
- Postal City and State
- Census 2010 Tabulation State, County, Tract and Block Code

    At a minimum, this information will allow us to update the source data for an existing address record, adding to our confidence that the address is valid

Address Data Content Guidelines 2

B) In order to update the location (geocode) for an existing MAF address, the submitted record must meet the requirements of "A" above, but this action requires either:

- Address Coordinate <u>or</u>
- Census 2010 Tabulation State, County, Tract and Block Code

This information will allow us to assign a location to previously ungeocoded addresses, to compare source geocodes to existing MAF geocodes, and to correct geocodes on MAF records where necessary

C) In order to ADD new records to the MAF, the submitted record must meet the requirements of "A" above, and additionally must include an Address Feature Type indicator identifying the address as residential, commercial, utility, etc.

Address Data Content Guidelines 3

D) In addition to the above, it is highly desirable that, where applicable, every address must have a unique <u>Within Structure Identifier</u> (Sub-address Element, for example Apt 3 or Unit A) to distinguish it from other addresses on the list. In the event that the specific unit designations are unknown, partners should supply:

- The number of units at the basic street address (i.e. "123 Main Street" "25 Units"). A basic street address (BSA) is a city style address[1] without the unit identifier.
- A flag indicating which addresses are multi-unit structures (i.e. apartment or condominium buildings)

E) For <u>Group Quarters</u>, source data should specify the NAME (i.e. Shady Acres Retirement Home) of the Group Quarters, and the TYPE (i.e. Hospital, Prison, College Dormitory) of Group Quarters represented by the address record.

A group quarters is a place where people live or stay that is normally owned or managed by an entity or organization providing housing and/or services for the residents. These services may include custodial or medical care as well as other types of assistance, and residency is commonly restricted to those receiving these services. People living in group quarters are usually not related to each other. Group quarters include such places as college residence halls, residential treatment centers, skilled nursing facilities, group homes, military barracks, correctional facilities, workers dormitories, and facilities for people experiencing homelessness.

F) The Census Bureau currently does not attempt to match or add to the MAF any address records that contain <u>only</u> Non-City-Style Addresses, such as:

- Rural Route Addresses, Post Office Box Addresses, Highway Contract Addresses, General Delivery, Location Descriptions (i.e. Brick House at intersection of 1st and Main Streets),
- Address Coordinates ONLY

Evaluation Stage

Geographers perform an interactive evaluation of a random sample and spatial clusters of the following categories:

Matched – No Differences

Matched – Geocoded

Matched with Differences

Unmatched

Evaluation Category 1
Matched – No Differences

- Validates/adds confidence to MAF Data – will update MAF Source and Quality Indicator (QI) scores

- If the partner structure coordinates are of good quality (and most are!), will add to the MTDB

United States™
Census
Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU

12

Evaluation Category 2 Matched – Newly Geocoded

- New address records from the USPS Delivery Sequence File may not be geocoded (missing TIGER Address Range)

- Where we match one of these to a local record with structure coordinates, the record becomes geocoded (~25% of the total in test areas)

Evaluation Category 3 Matched with Differences

- Geocode (Block) differences – partner coordinates place the address in a different block than the currently assigned block

- Feature Class (Address Type) – Census has the address classified as non-residential, partner data indicates a residence

United States™
Census
Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU

14

Evaluation Category 4 Unmatched

- Must evaluate a sample to determine if unmatched records indicate a true coverage issue – often indicative of different street names or other issues precluding a match (location vs. mailing addresses)

- Where appropriate, Add (obviously exists) or Provisional Add (likely to exist) new addresses and coordinates to the MTDB

Processing Results

For Two States and 5 Counties…

Total Records	1,614,597	
Duplicate Records	45,797	
Total Unduplicated	1,568,800	100 %
Matched to MTDB	1,448,464	92.3%
Unmatched	120,336	7.7%
Unmatched Residential	45,794	2.84%
Newly Geocoded	10,984	0.8%
Same Block	1,387,142	95.8%
Different Block	50,015	3.5%
Same Class	1,268,459	87.6%
Different Class	27,111	1.9%

~ 10% of the total records had no address classification code.

Positive Outcomes, Phase 1

- Improvements to ETL process
- Improvements to matching algorithms
- Highlights "problem areas" in the MAF
- Lessens impact of ungeocoded addresses
- Greater understanding of partner data
- Greater confidence in MAF data

Stuart Irby
301.763.9180
stuart.c.irby@census.gov